

Introduction to Value-Added Modeling for Accountability and Evaluation

Stefani Barbero
Manager of EVAAS Educator Support Services
Stefani.Barbero@sas.com
www.sas.com/evaas



THE
POWER
TO KNOW.

Why are value-added and growth models used today?

- Value-added/growth measures compare students' entering achievement and tested achievement to assess whether progress was adequate relative to some standard or expectation.
- This can be done in a way that allows a student to serve as their own control if it can take into account prior achievement adequately.
- Value-added and growth models have many uses. Some possibilities are:
 - Targeted professional development for educators
 - Identify educators' strengths and areas for improvement
 - Evaluate program/policy effectiveness and ROI
 - Provide more reliable data about the general achievement of students
 - Maximize potential educational opportunity for students of all levels of achievement
 - Component of State accountability measures at all levels of schooling influence
 - Federal accountability growth models

Value-added and growth models? Are they different?

Basically, it comes down to the intended use of the model.

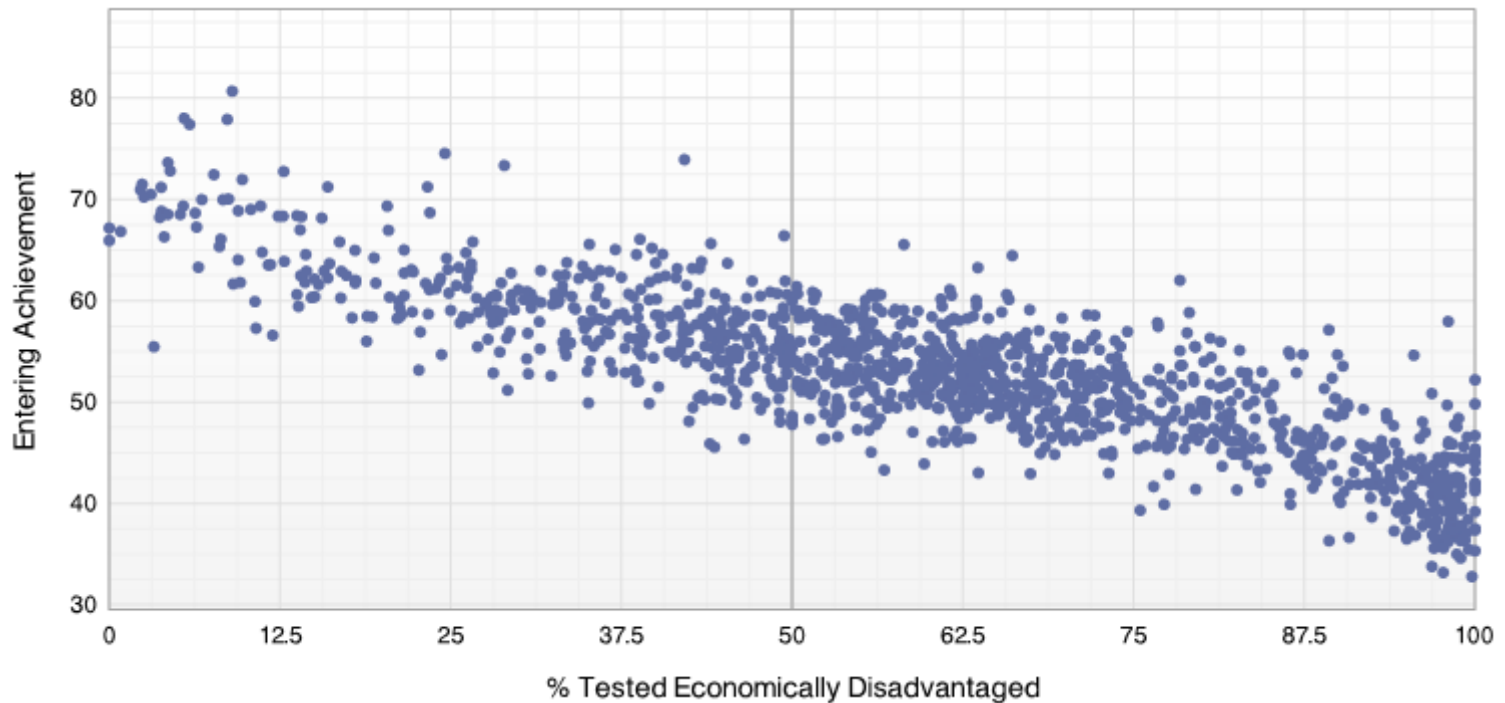
- If the focus is to simply measure student academic growth, then a growth model is appropriate
- If the focus is to measure student academic growth for a particular district/school/teacher then a value-added model is appropriate
- When these models are used for classification, they must be more rigorous

Characteristics of reliable value-added models

- While the concept of value-added is simple, the application is more complicated.
- Any value-added model must address the following inherent challenges in a statistically reliable approach:
 - Measurement error
 - Missing data
 - Different types of testing
 - Testing changes over time
 - Different student entering achievement levels
 - Student/teacher mobility
 - Shared instructional practices
- Simple growth models/VAMs do not address all of these issues

EVAAS in practice: student achievement & relationship to socioeconomic status

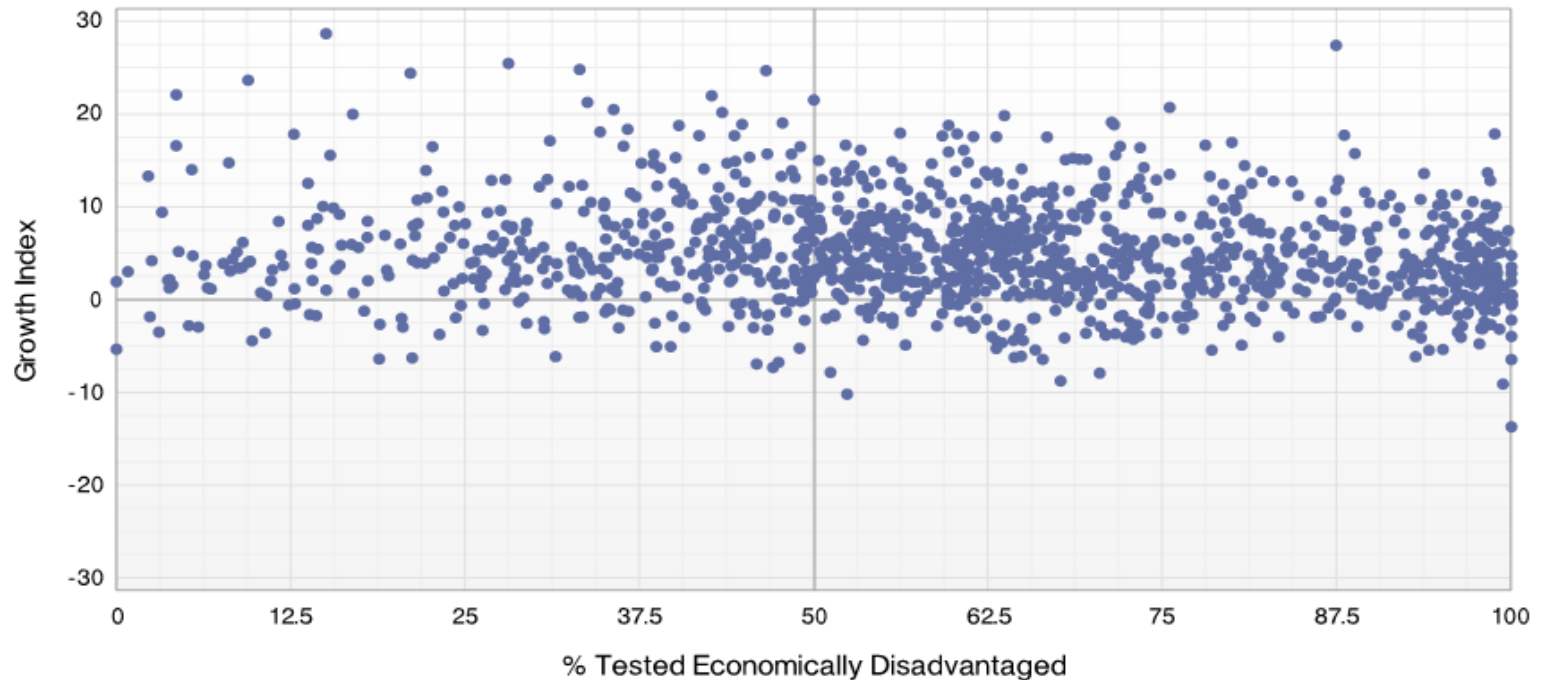
- It is well known that student achievement is often highly correlated to their socioeconomic or demographic characteristics.
- Evaluation and accountability models that only look at achievement will disadvantage educators serving low-achieving/high-poverty students.



Source: TVAAS Public Site; each dot represents a school.

EVAAS in practice: student progress & relationship to socioeconomic status

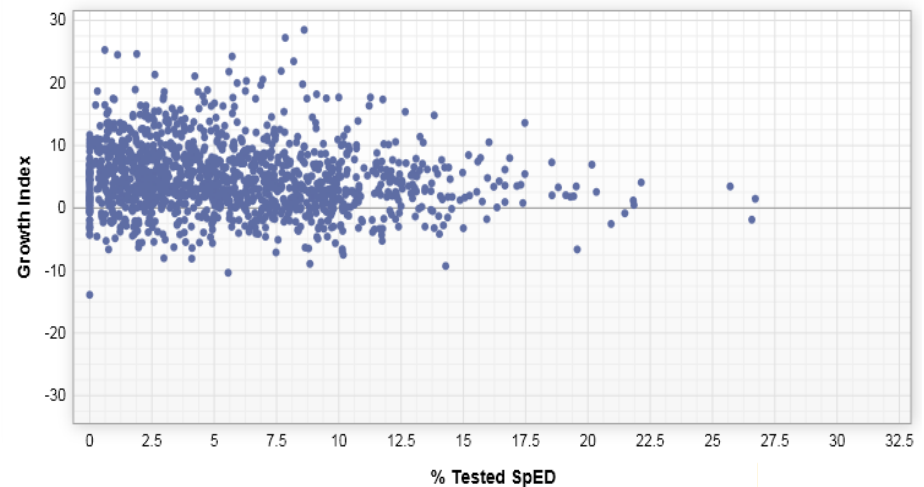
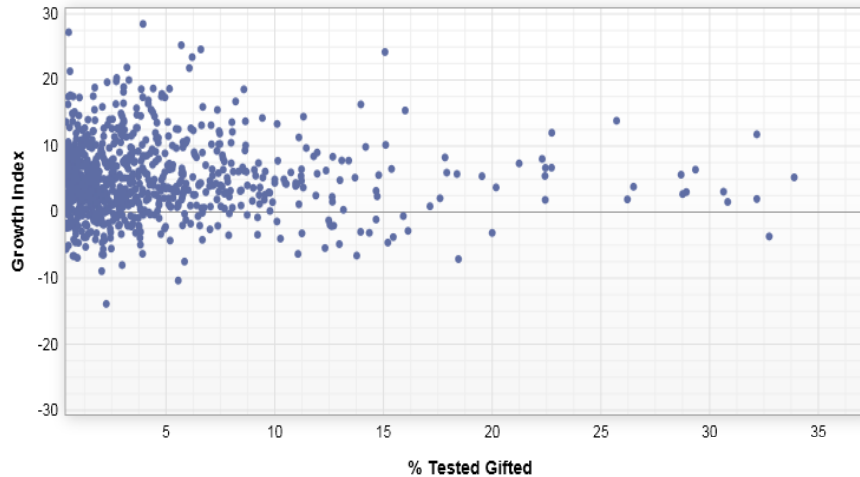
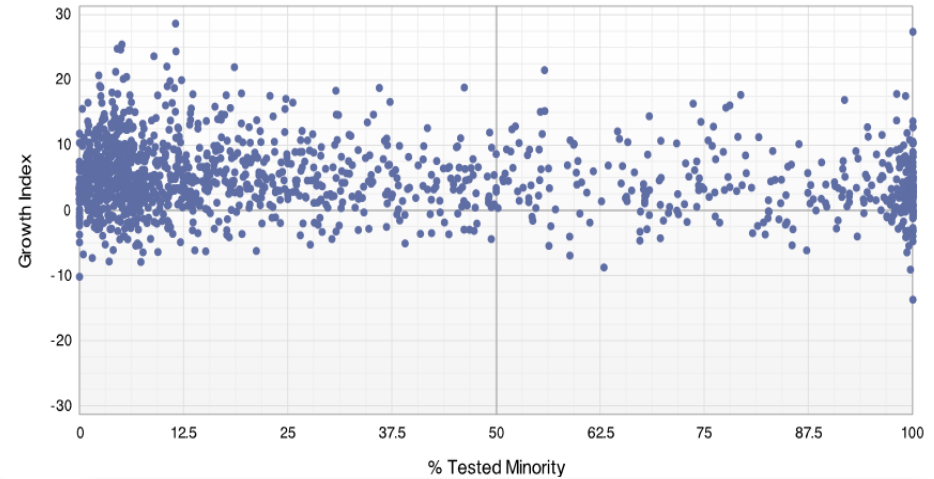
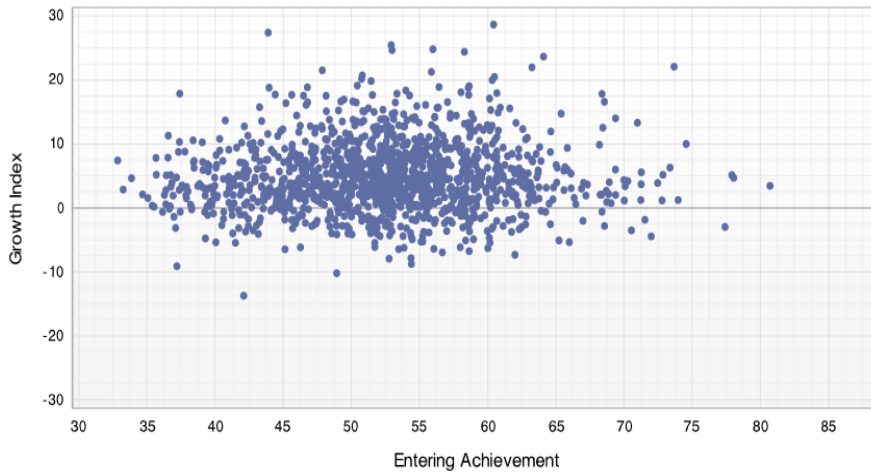
- Student progress (as measured by EVAAS) is not correlated to student characteristics, which is a more fair approach to evaluation/accountability.
- Educators can be shown to be effective regardless of the socioeconomic circumstances of their students.



Source: TVAAS Public Site; each dot represents a school.

EVAAS in practice: student progress

- Progress is equitable across achievement levels, demographics and for exceptional learners.



Source: TVAAS Public Site; each dot represents a school.

Student-Teacher Linkages

- SAS' roster verification system allows teachers to adjust the percentage of instructional time spent with each student prior to the value-added analysis.
- "The linchpin of efforts to improve teacher effectiveness is that **states must reliably link students and teachers** in ways that capture the complex connections that exist in schools," ~Data Quality Campaign fact sheet *Understanding Teacher Effectiveness*

Educators can accurately attribute:

- Student/teacher mobility
- Team teaching
- Push-in or Pull-out teachers
- Assistant teachers
- Long-term substitute teachers
- Contributing professionals
- Teacher assignment change
- Teacher turnover
- Intervention
- Tutoring
- Program participation
- Blended learning

A primary concern with more simplistic models: Repeatability/Stability

- Not all VAMs are created equal
- Single year vs. multiple year estimates
- Most year-to-year value-added teacher correlations range between 0.33 to 0.40
- SAS' methodology yields repeatability estimates between 0.70 to 0.80 for 3-year teacher estimates (based on 14 years of teacher-level reporting in one state)
- Highly effective teachers are very likely to remain effective.
- Less effective teachers may improve over time.
- If policymakers make decisions based on 3-year estimates, there is very little risk that the teachers identified as effective will be identified as ineffective three years later.

How does the state/district define growth?

- Sophisticated VAMs allow several options, such as the average progress made by students that year, in a reference year, etc.
- Some VAMs/growth models will only allow a “high half” and “low half” of teacher value-added/growth estimates each year,
- Not having this restriction and allowing all educators to demonstrate effectiveness can be crucial in accountability.
- Consider how you would like to define growth and what options you have for each model/approach.

How does the state/district define effectiveness?

- Can use the reference point, value-added/growth estimates and measures of uncertainty
- Can distinguish estimates into groups with equal proportions
- Can set risk differently for “above” and “below” categories

One System in practice: SAS® EVAAS®

- SAS has provided a comprehensive system including value-added reporting to educators for 20 years across 20 states.
- From that experience, we have learned:
 - Reliability of reporting is the most important factor in selecting a model.
 - Modeling flexibility is important so that policymakers have options in determining the growth expectation, levels of effectiveness, and can make different comparisons.
 - It is important to use multiple models if you do not want to force a “high half” and “low half” of effective educators.
 - Multiple year averages are key for accountability models.

Educators need more than a single estimate

- Need to provide a comprehensive continuous improvement system
- Web Interface with role-based security for permission and access
- Student-Teacher Linkages allowing roster verification at all levels
- District/school/teacher level *reflective* and diagnostic reporting on growth
 - Within subjects and grades
 - Within student achievement and demographic subgroups
- District/school/teacher and student level *proactive* reports that offer projections to future success on various academic milestones
 - Provide earlier RTI
 - Strategic placement of students into different courses
- Exporting of information for use in other data systems
- Educator support and training

EVAAS in practice: sample school report

2011 School Value Added Report Blue Middle School in Large Urban School District TCAP Math

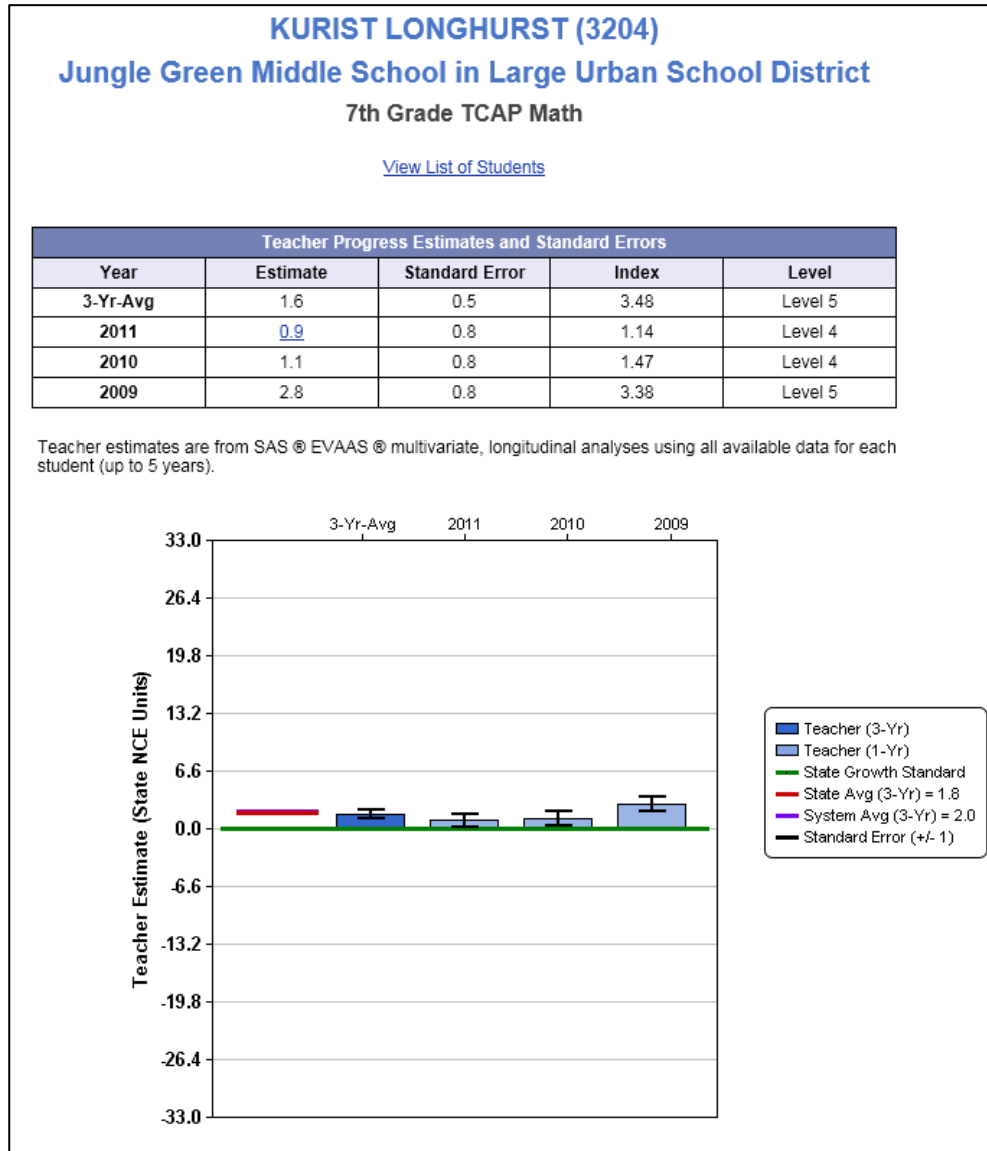
Estimated School Mean NCE Gain						
Grade	6	7	8	Mean NCE Gain over Grades Relative to		
Growth Standard	0.0	0.0	0.0			
State 3-Yr-Avg	0.1	1.8	1.5	Growth Standard	State	
2009 Mean NCE Gain	-3.8 R*	2.1 G*	1.6 G*	-0.1	-1.2	
Std Error	0.9	0.9	0.9	0.5	0.5	
2010 Mean NCE Gain	2.4 G*	3.9 G*	4.8 G*	3.7	2.6	
Std Error	0.9	0.9	0.9	0.5	0.5	
2011 Mean NCE Gain	4.1 G*	1.4 G*	4.5 G*	3.4	2.2	
Std Error	0.9	0.9	0.9	0.5	0.5	
3-Yr-Avg NCE Gain	<u>0.9 G*</u>	<u>2.5 G*</u>	<u>3.6 G*</u>	2.3	1.2	
Std Error	0.5	0.5	0.5	0.2	0.2	

Estimated School Mean NCE Scores				
Grade	6	7	8	
State Base Year (2009)	50.0	50.0	50.0	
State 3-Yr-Avg	43.7	43.5	43.6	
2008 Mean	24.9	26.8	22.6	
2009 Mean	28.9	27.0	28.4	
2010 Mean	35.9	32.8	31.9	
2011 Mean	41.9	37.4	37.3	

G*	Estimated mean NCE gain is above the growth standard by at least 1 standard error.
G	Estimated mean NCE gain is equal to or greater than growth standard but by less than 1 standard error.
Y	Estimated mean NCE gain is below the growth standard by 1 standard error or less.
R	Estimated mean NCE gain is more than 1 standard error below the growth standard but by 2 standard errors or less.
R*	Estimated mean NCE gain is below the growth standard by more than 2 standard errors.

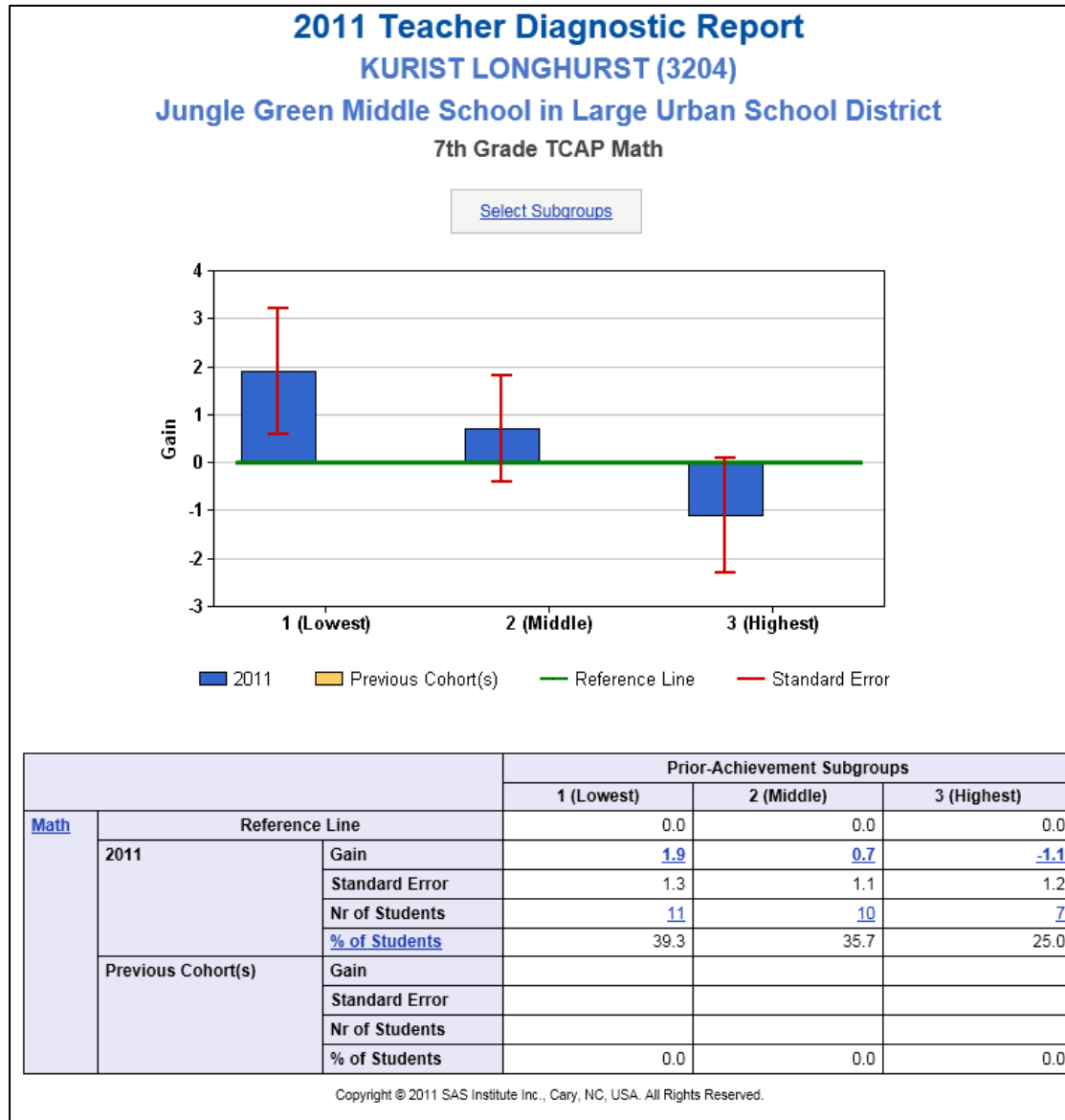
To view additional reports, click on the underlined numbers or words.

EVAAS in practice: sample teacher report



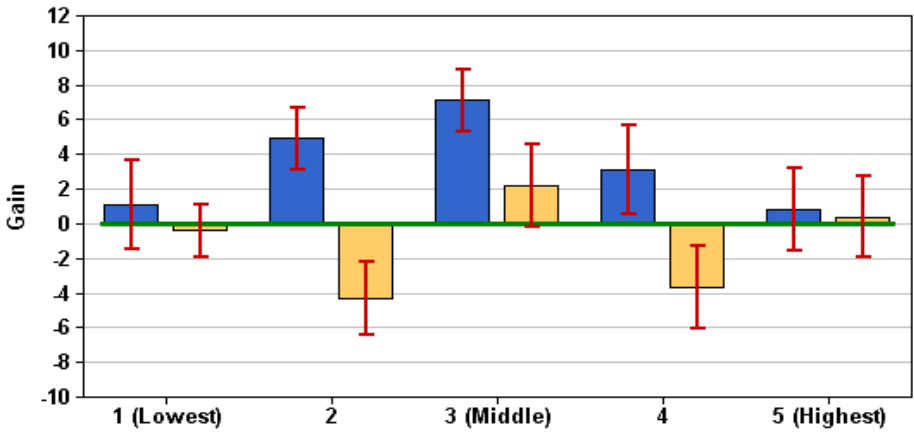
Copyright © 2011 SAS Institute Inc., Cary, NC, USA. All Rights Reserved.

EVAAS in practice: sample diagnostic report



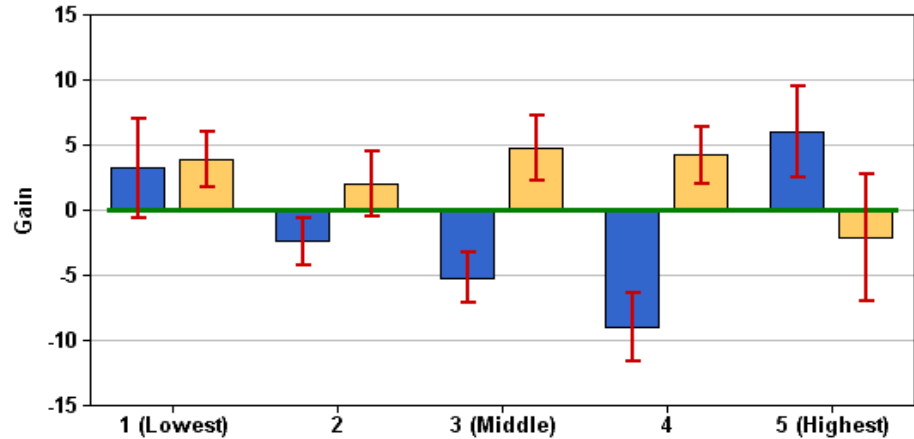
Report: School Diagnostic Test: PASS
 School: School-XXv Subject: Math
 District: District-DRw Grade: 4th Grade
 Year: 2012

Select Subgroups



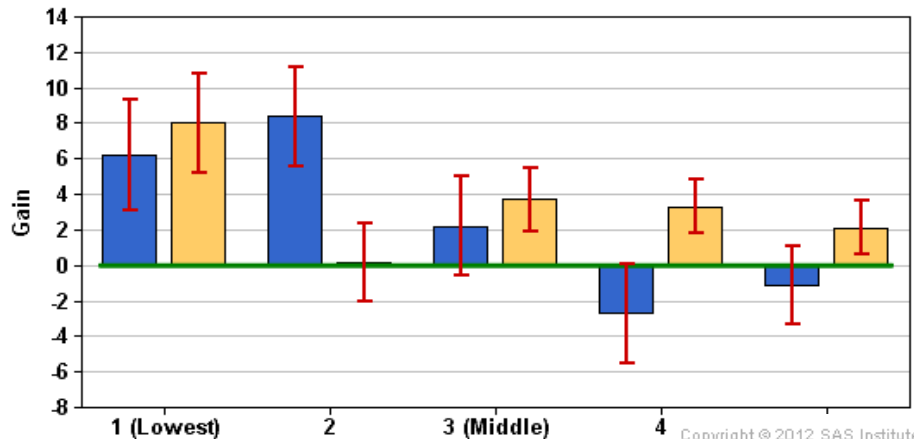
Report: School Diagnostic Test: PASS
 School: School-XXg Subject: Math
 District: District-DRw Grade: 4th Grade
 Year: 2012

Select Subgroups



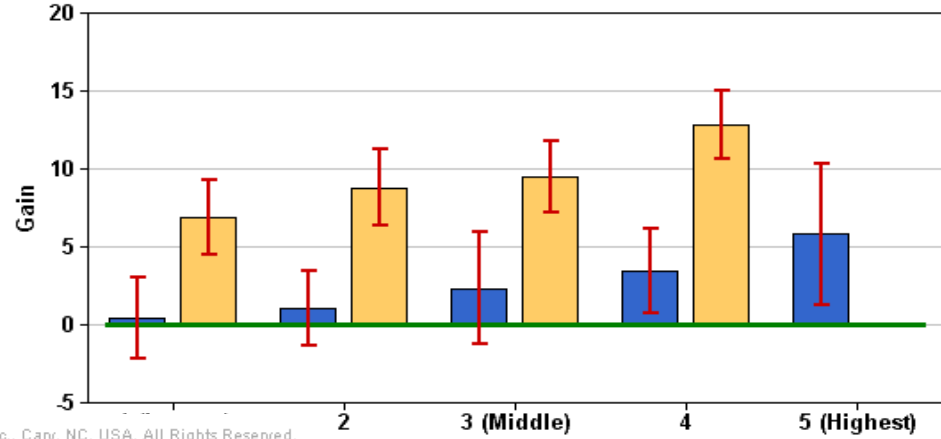
Report: School Diagnostic Test: PASS
 School: School-XXj Subject: Math
 District: District-DRw Grade: 4th Grade
 Year: 2012

Select Subgroups



Report: School Diagnostic Test: PASS
 School: School-FmS Subject: Math
 District: District-DRw Grade: 5th Grade
 Year: 2012

Select Subgroups



Looking Forward- Transitioning to Next Generation Assessments: PARCC and SBAC

- The SAS team behind EVAAS has gone through several major changes in assessments in many states in the past 20 years.
- EVAAS benefits from a flexible approach, which allows the use of tests, even when they are on differing scales or even when testing regimes have changed over time.
- Recommend setting an intra-year growth standard the year before and in year one of any new test implementation.
- Value-added results can then be provided in the first year of the new assessments.

Implementation of a value-added model in a comprehensive system

- One of the most important decisions is selecting the statistical approaches, since the results would be meaningless if they were biased and/or misleading.
- The decisions to ask yourself include:
 1. What type of value-added/growth model should be used?
 2. How will growth be defined?
 3. What reference group will be used for comparisons?
 4. What types of tests will be used, and do they meet some standard criteria?
 5. How will this information be reported to educators and administrators?
 6. What types of support will be available for educators?

A well designed value-added model

- Accommodates all of the challenges in testing
- Has high stability/repeatability
- Offers flexibility to policymakers and educators in determining the expectation of growth and defining effectiveness.
- Provides results that are not correlated with students' entering achievement or background so that educators are not advantaged or disadvantaged by the type of students they serve
- Can use CRTs, NRTs, college readiness tests, PARCC/SMARTER Balanced, or any other types of assessments that meet some basic criteria
- Takes into account estimation error and provides confidence intervals



John White, Ph.D.
Senior Manager of EVAAS Research and Development